FUNDAÇÃO
GETULIO VARGAS

EPGE
Escola de Pós-Graduação
em Economia

Ensaios Econômicos

Escola de
Pós-Graduação
em Economia
da Fundação
Getulio Vargas

N° 347 ISSN 0104-8910

Estimating and Forecasting the Volatility of Brazilian Finance Series Using Arch Models (Preliminary Version)

João Victor Issler

Junho de 1999

URL: http://hdl.handle.net/10438/577

Victor Issler, João
Estimating and Forecasting the Volatility of Brazilian Finance Series Using Arch Models (Preliminary Version)/ João Victor Issler - Rio de Janeiro : FGV,EPGE, 2010
(Ensaios Econômicos; 347)

Inclui bibliografia.

CDD-330

Estimating and Forecasting the Volatility of Brazilian Finance Series Using ARCH Models[¤]

João Victor Issler
Graduate School of Economics - EPGE
Getulio Vargas Foundation
Praia de Botafogo 190 s. 1125
Rio de Janeiro, RJ 22253-900
Brazil

Preliminary version, please do not quote. March, 1999

Abstract

The goal of this paper is to present a comprehensive emprical analysis of the return and conditional variance of four Brazilian …nancial series using models of the ARCH class. Selected models are then compared regarding forecasting accuracy and goodness-of-…t statistics. To help understanding the empirical results, a self-contained theoretical discussion of ARCH models is also presented in such a way that it is useful for the applied researcher. Empirical results show that although all series share ARCH and are leptokurtic relative to the Normal, the return on the US$ has clearly regime switching and no asymmetry for the variance, the return on COCOA has no asymmetry,

[¤]This paper was prepared for the Invited Session on volatility of the Sociedade Brasileira de Econometria (SBE) conference in Vitória, 1998. I thank the Editors of Revista de Econometria for the invitation to participate in this Session, Ajax Moreira and seminar participants for helpful comments and suggestions, and Jim Hamilton for providing the GAUSS code to run SWARCH models. Discussions with Farshid Vahid were very helpful for implementing SWARCH-model estimation, and are gratefully acknowledged. All remaining errors are mine. Financial support from CNPq-Brazil and PRONEX are gratefully acknowledged.

while the returns on the CBOND and TELEBRAS have clear signs of asymmetry favoring the leverage effect. Regarding forecasting, the best model overall was the EGARCH(1,1) in its Gaussian version. Regarding goodness-of-fit statistics, the SWARCH model did well, followed closely by the Student-t GARCH(1,1).

1 Introduction

ARCH - Autoregressive Conditional Heteroskedasticity - is recognized today as a major feature of financial data which several econometric models try to capture. This paper presents a comprehensive empirical analysis of the return and conditional variance of a variety of Brazilian financial series using models of the ARCH class. The data used covers a wide spectrum of Brazilian Finance series: a spot stock-price index - IBOVESPA, of the São Paulo Stock Exchange, the spot price of a popular Brazilian stock - TELEBRAS, traded in the São Paulo Stock Exchange, a spot currency-exchange rate - R$/US$, and a spot popular commodity price - COCOA. Selected models of the ARCH class for the return of these series are fitted and estimates compared regarding forecasting accuracy and goodness-of-fit statistics. To help to understand the empirical results obtained, a self-contained theoretical discussion of ARCH models is also presented, focusing on econometric results of ARCH models that are useful for the applied researcher.

Empirical results show that all series share ARCH and are leptokurtic relative to the Normal. However, they all have specificities as well: the return on the US$ has regime switching and no asymmetry for the variance, the return on COCOA has no asymmetry, while the returns on the CBOND and TELEBRAS have signs of asymmetry favoring the leverage effect. This stresses the point that successful modelling of asset returns requires taking into account the essential features of each of them separately. Regarding forecasting accuracy, the best model overall was the EGARCH(1,1) in its Gaussian version. This may be due to the fact that, for Brazilian data, outliers are a common occurrence. Regarding goodness-of-fit statistics, the SWARCH model did well, followed closely by the Student-t GARCH(1,1).

2 Some Theory of ARCH Models

In this Section selected results for ARCH models are presented. Since the literature on ARCH is vast and comprehensive, including more than one hundred papers and the surveys of Bollerslev, Chou and Kroner(1992), of Bollerslev, Engle and Nelson(1994), and the book of collected papers edited by Engle(1995), it makes little sense to repeat here theoretical results that are already discussed elsewhere at greater depth and length.

Instead, the focus of this Section is on econometric results of ARCH models that are useful for the applied researcher. In a direct analogy with the time-series method proposed by Box and Jenkins(1976), the "identification" of ARCH processes is motivated here by considering autocorrelation and partial autocorrelation functions for the squared (returns of financial) series. We later discuss three improvements for the class of ARCH models: the generalized ARCH - GARCH, the Exponential GARCH - EGARCH, and the switching-regime ARCH model - SWARCH, motivating their introduction by previous shortcomings in applying early ARCH models to actual (Finance) data. All models discussed here are later applied to Brazilian financial data.

2.1 ARCH: The Basic Idea and Selected Models

Time series $\{Y_t\}_{t=1}^{\infty}$ have two basic properties that made modelling and forecasting them easy to a wide audience; see Box e Jenkins(1976) and all the literature that follows. The first is weak stationarity, which imposes the restriction that $E[Y_t] = \mu, \forall t$, and $E[(Y_t - \mu)(Y_{t-j} - \mu)] = \gamma_j, \forall t, j$. These restrictions may apply to the original series $\{Y_t\}_{t=1}^{\infty}$ or to some transformation of it (e.g., its first differences). Since time only goes forward, the econometrician has to form time averages to estimate consistently population parameters of time-series models. Thus, without this property, it would be impossible to conceive consistent estimates due to the lack of degrees of freedom. The second property is autocorrelation, i.e., the fact that it is usually the case that $\gamma_j \neq 0$ for some $j > 0$. Again, this property may apply to the original series $\{Y_t\}_{t=1}^{\infty}$ or to a transformation of it (e.g., its square). Autocorrelation allows modelling time series based on their own "past," and on an unpredictable error (ε_t) and its own "past." Thus, forecasts will be a function of current and past values of Y_t and ε_t. In models popularized by Box and Jenkins, the functional form is linear, but there are several examples of models that use a non-linear function. Perhaps, the most important one in Finance is the class of ARCH models introduced by Engle(1982).

Although the ARCH model and its extensions are widely applied in Finance today, that was not at all its original motivation. Engle(1995, pp. xi-xii) writes that he thought that the main contribution of ARCH models would be on the Rational Expectations debate in Macroeconometrics. By the end of the 1970's, inflation was soaring everywhere due to two oil shocks. This raised concerns of how well it could be forecast. Indeed, Okun(1971) and Friedman(1977) proposed that an increase in the level of inflation would raise its variance. Interestingly enough, using the framework of his 1982 Econometrica paper, Engle(1983) showed that the high level of U.S. inflation experienced around that time was highly predictable, and that inflation level and variance were uncorrelated.

As is the case with several successful theoretical developments in econometrics, Clive Granger reports that the birth of ARCH models was a consequence of an observed empirical regularity. When applying the tools proposed by Box and Jenkins, one could find series which were unpredictable, although their squares were highly predictable. This suggested that some sort of nonlinearity was at work. This lead Engle(1982) to propose the class of ARCH models, which conforms to this early empirical observation. Granger, on the other hand, went on to work with the bilinear model, which has a structure similar to that of ARCH models; see Granger and Andersen(1978).

The linear ARCH(p) model introduced by Engle can be summarized in:

$$\begin{aligned} y_t &= x_t \beta + \varepsilon_t, \\ \varepsilon_t | \Omega_{t-1} &\sim D\left(0; \sigma_t^2\right), \\ \sigma_t^2 &\equiv E\left[\varepsilon_t^2 | \Omega_{t-1}\right] = \alpha_0 + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2, \end{aligned} \tag{1}$$

where $D(\cdot)$ is some parametric distribution, usually the Normal or the Student-t, and x_t is either weakly exogenous or an element of the conditioning set Ω_{t-1}. Notice the similarity between the last line in (1) and an AR(p) model:

$$W_t = c + \sum_{i=1}^{p} \phi_i W_{t-i} + \mu_t. \tag{2}$$

It is easy to write an ARCH process (1) as an autoregressive process (2), using the result that $\varepsilon_t^2 = E[\varepsilon_t^2 | \Omega_{t-1}] + \eta_t$, where $\{\eta_t\}_{t=1}^{\infty}$ is a Martingale-

Difference Sequence. We have:

$$\varepsilon_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_p \varepsilon_{t-p}^2 + \eta_t. \tag{3}$$

To ensure that $\varepsilon_t^2 \geq 0, \forall t$, for all possible realizations of $\{\varepsilon_t\}_{t=1}^{\infty}$, it is usually assumed that (i) $\alpha_0 > 0$ and $\alpha_i \geq 0, \forall i = 1; \cdots; p$, and that (ii) η_t has a lower bound of $-\alpha_0$.

The simplest linear ARCH model possible - the Gaussian ARCH(1), with $\beta = 0$, allows discussing several interesting features of models in this class. It is summarized in:

$$\begin{aligned} Y_t | \psi_{t-1} &\sim N\left(0; \sigma_t^2\right) \\ \sigma_t^2 &= \alpha_0 + \alpha_1 Y_{t-1}^2. \end{aligned} \tag{4}$$

Notice that if $\alpha_1 = 0$, $Y_t | \psi_{t-1} \sim N(0; \alpha_0)$, i.e., Y_t is conditionally homocedastic. The theorem below shows some of its basic properties.

Theorem 1 (Engle(1982)) For integer r, the $2r$th (unconditional) moment of a first-order linear ARCH process with $\alpha_0 > 0$, $\alpha_1 \geq 0$ exists if, and only if, $\alpha_1^r \prod_{j=1}^{r} (2j-1) < 1$.

Thus, the second and fourth unconditional moments are defined, with $E(Y_t^2) = \frac{\alpha_0}{1-\alpha_1}$, and $E(Y_t^4) = \frac{3\alpha_0^2}{(1-\alpha_1)^2} \cdot \frac{1-\alpha_1^2}{1-3\alpha_1^2}$, iff $\alpha_1 < 1$ and $3\alpha_1^2 < 1$ respectively. This has two implications. First, although the conditional model is Gaussian, the unconditional distribution has fatter tails compared to the Normal[1]. This happens because $\frac{1-\alpha_1^2}{1-3\alpha_1^2} > 1$. Under a Gaussian unconditional distribution, the Kurtosis coefficient is given by $3\frac{\alpha_0^2}{(1-\alpha_1)^2}$, hence smaller than $\frac{3\alpha_0^2}{(1-\alpha_1)^2} \cdot \frac{1-\alpha_1^2}{1-3\alpha_1^2}$. Second, although the conditional distribution is heteroskedastic, the conditional distribution is homocedastic. Indeed, because the mean and autocovariances of Y_t are not time varying, this process is weakly stationary despite displaying conditional heteroskedasticity. This result generalizes for a wider class of ARCH(p) models as below.

[1] Despite the fact that this feature of conditional normality is helpful in modelling fat-tailed distributions, it is definitely not enough to account for the observed pattern of leptokurtosis on asset returns. The alternative of using the Student-t distribution or the Generalized Error Distribution is discussed below. Bollerslev, Engle and Nelson(1994) also discuss the use of the Generalized-t distribution.

Theorem 2 (Engle(1982)) The pth-order linear ARCH process with $\alpha_0 > 0$, $\alpha_1, \cdots, \alpha_p \geq 0$ is covariance- (weakly-) stationary if, and only if, the associated characteristic equation has all roots outside the unit circle. The stationary variance is given by $\frac{\alpha_0}{1-\sum_{i=1}^{p}\alpha_i}$.

This last result illustrates why the appropriate ARCH (p) model conforms to the stylized fact that the level series (error) is white noise despite the fact that the autocorrelation and partial autocorrelation of its squares show signs of predictability. Following Bollerslev(1986), write the error term in (3) as $\varepsilon_t = \sigma_t \cdot z_t$, where $z_t \sim i.i.d.(0,1)$. Based on the law of iterated expectations is easy to show that $E[\varepsilon_t] = 0$, $E[\varepsilon_t \varepsilon_{t-j}] = 0 \; \forall j > 0$, and given Theorem 2, $E[\varepsilon_t^2] = \frac{\alpha_0}{1-\sum_{i=1}^{p}\alpha_i}$. Hence, ε_t is white noise. However, if we take $\varepsilon_t^2 = \sigma_t^2 \cdot z_t^2$, it is obvious that its autocorrelations will not die out because of (3), and that its partial autocorrelation will only be zero starting at order $p+1$.

Estimating an ARCH (p) process by maximum likelihood is straightforward once a parametric distribution is assumed for z_t in $\varepsilon_t = \sigma_t \cdot z_t$. The usual assumption is that z_t has an $i.i.d.$ Gaussian or Student-t distribution. In any case, the joint density of the sample $y_1, \cdots, y_T$ can be recursively factored into the conditional and marginal densities to form:

$$f(y_1, \cdots, y_T; \theta) = \prod_{t=1}^{T} f(y_t | y_{t-1}, \cdots, y_{-p+1}; \theta); \tag{5}$$

where θ is a vector of parameters of the joint density, and we condition on pre-sample observations. For the simple ARCH (1) model discussed above we have:

$$f(y_t | y_{t-1}, \cdots, y_0; \theta) = \frac{1}{\sqrt{2\pi} \cdot \left(\alpha_0 + \alpha_1 y_{t-1}^2\right)^{1/2}} \times \exp\left(-\frac{1}{2}\left(\frac{y_t^2}{\alpha_0 + \alpha_1 y_{t-1}^2}\right)\right); \tag{6}$$

where $\theta = (\alpha_0, \alpha_1)'$. Using this result, maximum likelihood estimates can be found by numerically optimizing the conditional log-likelihood function:

$$\log L(\theta; \cdot) = T \cdot \log\left(\frac{1}{\sqrt{2\pi}}\right) - \frac{1}{2}\sum_{t=1}\left(\log\left(\alpha_0 + \alpha_1 y_{t-1}^2\right) + \frac{y_t^2}{\alpha_0 + \alpha_1 y_{t-1}^2}\right); \tag{7}$$

subject to the constraints $\alpha_0 > 0$ and $\alpha_1 \geq 0$.

For Financial series, if one believes that outliers are clustered (e.g., Mandelbrot(1963)), then the estimation method described above reduces the impact of these extreme observations on parameter estimates. This happens because the denominator of $\frac{y_t^2}{\alpha_0+\alpha_1 y_{t-1}^2}$ will reduce the contribution of a given outlier y_t^2 in (7), since y_{t-1}^2 is more likely to be large as well. This shows a clear advantage for recognizing the presence of ARCH when compared to (say) the case where homocedasticity is erroneously assumed. In the latter, all observations are equally weighted in the likelihood function, whereas in the former outliers get a smaller weight. Coupled with the fact that the unconditional distribution is leptokurtic relative to the Normal, allowed Engle(1982) to conclude that ARCH models show potential to deal with clustered outliers. Of course, if outliers are not clustered, the weighting procedure will not work. Hence, Engle mentions nothing regarding outliers in general. He also makes no attempt to compare ARCH models with "robust" estimates[2].

From empirical experimentation with models in the ARCH(p) class, it became apparent that the order of the fitted model was quite large - p large. In a direct analogy with models in the AR(p) class, where the parsimonious solution is to include MA$(\cdot)$ terms - forming an ARMA model, the ARCH(p) process was generalized to include these "MA(q) terms." This is the motivation behind the GARCH(p;q) model, proposed by Bollerslev(1986)[3]:

$$\begin{aligned}
\varepsilon_t | \psi_{t-1} &\sim D\left(0;\ \sigma_t^2\right); \\
\sigma_t^2 &= E\left[\varepsilon_t^2 | \psi_{t-1}\right] = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2 \\
&= \alpha_0 + A(L)\ \varepsilon_t + B(L)\ \sigma_t^2; \qquad (8)
\end{aligned}$$

where $A(L) = \sum_{i=1}^{q} \alpha_i L^i$ and $B(L) = \sum_{i=1}^{p} \beta_i L^i$ are finite order polynomials on the lag operator L. To see how the solution proposed by Bollerslev mimics the ARMA class, consider for simplicity the GARCH(1;1) model:

$$\sigma_t^2 = \omega + \alpha\, \varepsilon_{t-1}^2 + \beta\, \sigma_{t-1}^2 : \qquad (9)$$

If we define as in (3), $\eta_t \equiv \varepsilon_t^2 - \sigma_t^2$ as the conditional variance prediction error, with the property that $E[\eta_t | \psi_{t-1}] = 0$, we can solve (9) in terms of

[2] See the discussion in Engle(1982), at the end of Section 3.

[3] See Engle(1995, p. xii) for a historical account.

current and lagged ε_t^2 and η_t to get:

$$\varepsilon_t^2 = \omega + (\alpha + \beta)\, \varepsilon_{t-1}^2 - \beta\, \eta_{t-1} + \eta_t, \tag{10}$$

which is an $ARMA(1,1)$ process for ε_t^2. Using the same principle, it is easy to show that a $GARCH(p,q)$ model is indeed an $ARMA(\max(p,q), p)$ for ε_t^2; see Bollerslev(1986).

The Theorem below shows that the result in Theorem 2 generalizes for this wider class of ARCH models.

Theorem 3 (Bollerslev(1986)) *The Gaussian $GARCH(p,q)$ process, with $\omega > 0$, $\alpha_i, \beta_i \geq 0$, $\forall i = 1, \cdots, \max[p,q]$, is weakly stationary with $E(\varepsilon_t) = 0$; $Var(\varepsilon_t) = \frac{\omega}{(1 - A(1) - B(1))}$ and $Cov(\varepsilon_t, \varepsilon_s) = 0$, $t \neq s$; if, and only if, $A(1) + B(1) < 1$.*

The analogy between the ARMA and the GARCH class is also present when forecasting is considered, because the $GARCH(p,q)$ is linear on lagged ε_t^2 and σ_t^2. For the $GARCH(1,1)$ model, defining $E_t\sigma_{t+s}^2 \equiv E\left[\sigma_{t+s}^2 \mid \Omega_t\right]$ as the forecast of the conditional variance for horizon s, using information up to t, we have:

$$\begin{aligned} E_t\sigma_{t+2}^2 &= \omega + (\alpha + \beta)\, \sigma_{t+1}^2, \text{ and,} \\ E_{t+1}\sigma_{t+3}^2 &= \omega + (\alpha + \beta)\, \sigma_{t+2}^2. \end{aligned} \tag{11}$$

Taking the conditional expectation of the last line of (11) using the conditioning set Ω_t, it is easy to show using the law of iterated expectations that:

$$\begin{aligned} E_t\sigma_{t+3}^2 &= \omega\,[1 + (\alpha + \beta)] + (\alpha + \beta)^2\, \sigma_{t+1}^2 \\ &\vdots \\ E_t\sigma_{t+s}^2 &= \omega \left[\frac{1 - (\alpha + \beta)^{s-1}}{1 - (\alpha + \beta)}\right] + (\alpha + \beta)^{s-1}\sigma_{t+1}^2. \end{aligned} \tag{12}$$

As long as the parameter-restriction and stationarity conditions in Theorem 3 hold, i.e., $0 < \alpha + \beta < 1$, the expression in (12) converges to the unconditional variance of ε_t as the forecasting horizon increases:

$$E_t\sigma_{t+s}^2 \to \left(\frac{\omega}{1 - (\alpha + \beta)}\right), \; s \to \infty. \tag{13}$$

The Exponential GARCH - EGARCH model was proposed by Nelson(1991) to deal with three basic shortcomings of models in the GARCH class[4]. First, the impact of shocks on volatility is symmetric for these models. Hence, positive or negative shocks have exactly the same effect on the conditional variance. Since most applications of models in the GARCH class are in Finance, and for these data it is observed that the effects of positive and negative returns on volatility is not identical (e.g., Black(1976)), it is desirable to conceive models that allow estimation and testing for asymmetry. Second, the restrictions $\alpha_0 > 0$, $\alpha_i, \beta_i \geq 0, \forall i$, constrain the roots of the characteristic polynomials of GARCH models, preventing random oscillatory behavior in σ_t^2. Moreover, when these restrictions are binding, maximum likelihood estimates are a constrained optima. Third, measures of persistence of shocks to the conditional variance for integrated GARCH processes depend on the norm considered, and no direct analogy can be made with results of the unit-root literature[5].

We consider here a simpler version of the EGARCH model proposed by Nelson:

$$\begin{aligned}
\varepsilon_t &= \sigma_t \cdot z_t, \text{ with}, \\
z_t &\sim i.i.d.(0,1); \\
g(z_t) &= \theta z_t + \gamma\left[|z_t| - E(|z_t|)\right]; \\
\ln\left(\sigma_t^2\right) &= \omega + \left(\frac{1+\psi_1 L + \cdots + \psi_1 L^q}{1-\phi_1 L - \cdots - \phi_p L^p}\right) \cdot g(z_{t-1}) \\
&= \omega + \sum_{i=1}^{\infty} \beta_i g(z_{t-i}). \qquad (14)
\end{aligned}$$

The third line in (14) allows for asymmetric effects of shocks on the (log of the) conditional variance. When $z_t > 0$, the slope of $g(z_t)$ is $\theta + \gamma$, but for $z_t < 0$, the slope of $g(z_t)$ is $\theta - \gamma$. The fourth line depicts a simple ARMA process for $\ln(\sigma_t^2)$, proposed by Nelson as a parsimonious representation for the infinite MA process for $\ln(\sigma_t^2)$. In his original formulation, he also considered a time varying intercept for $\ln(\sigma_t^2)$.

Theorem 4 (light version of Nelson(1991)) If at least one of the parameters γ or θ are non-zero, $\{\exp(-\omega)\sigma_t^2\}$, $\{\exp(-\omega/2)\varepsilon_t\}$, and

[4] See the discussion in the Introduction of Nelson(1991).

[5] For this last point, see the discussion in Bollerslev, Engle and Nelson(1994, pp. 2990-2992).

$\{\ln(\sigma_t^2) - \omega\}$ are strictly (strongly) stationary and ergodic, and $\{\ln(\sigma_t^2) - \omega\}$ is covariance-stationary if, and only if, $\sum_{i=1}^{\infty} \beta_i^2 < \infty$.

It is worth noting that the ARMA specification presented above is widely used in applications, being the most relevant for the applied researcher. In this case, a necessary and sufficient condition for $\sum_{i=1}^{\infty} \beta_i^2 < \infty$ is that all the roots of $\left(1 - \phi_1 z - \cdots - \phi_p z^p\right) = 0$ lie outside the unit circle.

For the EGARCH $(1;1)$ model, the (log) variance equation is:

$$\ln\left(\sigma_t^2\right) = \omega' + \beta \ln\left(\sigma_{t-1}^2\right) + \alpha \left|\frac{\varepsilon_{t-1}}{\sigma_{t-1}}\right| + \gamma \frac{\varepsilon_{t-1}}{\sigma_{t-1}}; \tag{15}$$

where the model has been reparameterized with $\omega' = (1 - \phi_1)\,\omega - \gamma E\,|z_{t-1}|$, and $\beta = \phi_1$. A slightly different specification is considered in Hamilton(1994, pp. 668-669):

$$\left[\ln\left(\sigma_t^2\right) - \omega\right] = \beta \left[\ln\left(\sigma_{t-1}^2\right) - \omega\right] + \alpha \left(\left|\frac{\varepsilon_{t-1}}{\sigma_{t-1}}\right| - E\left|\frac{\varepsilon_{t-1}}{\sigma_{t-1}}\right| + \gamma \frac{\varepsilon_{t-1}}{\sigma_{t-1}}\right): \tag{16}$$

In either (15) or (16) there is no asymmetry in the variance as long as $\gamma = 0$. This constitutes a testing procedure for the asymmetric effect. If $\gamma \neq 0$, then there is a differentiated impact of news on volatility. If $\gamma < 0$ there is the so called "leverage effect," where good news have a smaller impact on the conditional variance than bad news; see Pagan and Schwert(1990) and Engle and Ng(1991).

Another model that allows estimating and testing the leverage effect is the Threshold ARCH - TARCH model, proposed independently by Zakoian(1990), and Glosten, Jaganathan, and Runkle(1993). The specification for the conditional variance is:

$$\sigma_t^2 = \omega + \beta \sigma_{t-1}^2 + \alpha \varepsilon_{t-1}^2 + \gamma \cdot d_{t-1} \cdot \varepsilon_{t-1}^2; \tag{17}$$

where the dummy variable $d_{t-1} = 1$, if $\varepsilon_{t-1} < 0$, and $d_{t-1} = 0$ otherwise. Again, there is no asymmetry in the variance as long as $\gamma = 0$. Here, there is the leverage effect if $\gamma > 0$.

At least since Bollerslev(1986), there was a switch in focus from Macroeconometrics to Finance for models within the ARCH class. By the time Nelson proposes the EGARCH model, this change is already consolidated. Later, the work of Hamilton and Susmel(1994) on ARCH and switching-regime models is motivated by the fact that "financial markets sometimes appear quite calm and other times highly volatile." According to Diebold(1986) and Lamoureux and Lastrapes(1990), one of the consequences of ignoring possible changes in regime is an overestimation of the persistence of shocks to the conditional variance. Hamilton and Susmel note that this relates "to Perron's(1989) observation that changes in regime may give the spurious impression of unit roots."

Nowhere there are more changes in rules and is volatility more variable than in emerging-market economies. Because this feature of emerging-market financial data leads to potentially interesting applications of techniques that recognize changes in regime, we discuss now the switching ARCH - SWARCH model of Hamilton and Susmel. The SWARCH $-$ L$(k; q)$ model, where the L stands for leverage (or asymmetric effect), k denotes the number of regimes, and q denotes the order of the ARCH process, is given by:

$$
\begin{aligned}
u_t &= \sqrt{g_{s_t}} \times \varepsilon_t; \\
\varepsilon_t &= \sigma_t \cdot z_t; \text{with}; \\
z_t &\sim i.i.d.(0; 1); \\
\sigma_t^2 &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_q \varepsilon_{t-q}^2 + \xi \cdot d_{t-1} \cdot \varepsilon_{t-1}^2;
\end{aligned}
\tag{18}
$$

where $d_{t-1} = 1$, if $\varepsilon_{t-1} \leq 0$, and $d_{t-1} = 0$, if $\varepsilon_{t-1} > 0$ is the dummy variable for the leverage effect, and s_t represents all the possible regimes for the variance process. For regime one, i.e., when $s_t = 1$, the variance factor is normalized to unity, i.e., $g_1 = 1$. When $s_t = 2$, all things equal, the variance of u_t is g_2 times higher than that in regime one, and so on, up to g_k, where $s_t = k$.

As in the rest of the literature on ARCH, it is usually assumed that z_t is either Gaussian of has a Student-t distribution. It is further assumed that s_t can be described by a Markov chain. The probability that there is a change in regime on t, going from regime i in $t - 1$, to regime j, is:

$$
\Pr[s_t = j \mid s_{t-1} = i] = p_{ij}; \tag{19}
$$

where it is useful to collect all these parameters in a transition probability matrix, $P = (p_{ij})$, of order $k \times k$. Notice that $\sum_{j=1}^{k} p_{ij} = 1$. Hence, the columns of P add up to unity. Along with the variance factors g_1, g_2, $\cdots$, g_k, these probabilities are additional parameters to be estimated. In forming the likelihood function, it is recognized that observations may come from any of these states, and thus probabilities are used as weights; see Hamilton and Susmel for details.

2.2 Estimation, Inference and Testing

It is common practice to estimate the models discussed above by maximum likelihood after a parametric distribution for the error term is assumed. For applied research, since there is usually doubt about which parametric distribution to use, it is helpful to regard these estimates as quasi-maximum likelihood; see Bollerslev and Wooldridge(1992) inter-alia. There is also the possibility of estimating ARCH processes by non-parametric, semi-parametric and semi-non-parametric methods; see Engle and Gonzalez-Rivera(1991) for semi-parametric methods, Hamilton(1994) for a discussion on non-parametric estimates of ARCH processes using the generalized method of moments (GMM), and Gallant and Tauchen(1989) and Gallant et al.(1991, 1992, 1993) for semi-non-parametric methods.

To perform conditional maximum likelihood estimation, first decompose the joint density of the sample $w_1, \cdots, w_T$ recursively as a product of conditional densities to form:

$$f(w_1; \cdots; w_T; \theta) = \prod_{t=1} f(w_t | w_{t-1}; \cdots; w_{-k}; \theta); \tag{20}$$

where θ is a vector of parameters of the joint density, w_t is a vector that includes the explained and explanatory variables, and conditioning on pre-sample observations up to $-k$ is implicit.

There are several examples of parametric densities that are used in practice in forming (20). Bollerslev(1986) assumes conditional normality for the error term ε_t of the GARCH$(p;q)$ model in (8), i.e.:

$$f(w_t | w_{t-1}; \cdots; w_{-k}; \theta) = \frac{1}{\sqrt{2\pi \cdot \sigma_t}} \times \exp\left(-\frac{1}{2}\frac{\varepsilon_t^2}{\sigma_t^2}\right); \tag{21}$$

whereas Bollerslev(1987) assumes that ε_t has a Sudent-t distribution with ν degrees of freedom and scale parameter M_t (to yield a unit variance). Hence,

$$f(w_t | w_{t-1}; \cdots; w_{-k}; \theta) = \frac{\Gamma[(\nu+1)/2]}{(\pi\nu)^{1/2}\,\Gamma[\nu/2]} M_t^{-1/2} \left(1 + \frac{\varepsilon_t}{M_t \nu}\right)^{-(\nu+1)/2}; \tag{22}$$

where $\Gamma[\cdot]$ is the gamma function.

The density in (22) allows for fatter tails than the one in (21). The parameter ν controls how fat the tails are. Since the Normal can be thought as a limiting case of the Student-t density when $\nu \to \infty$, using (22) in place of (21) allows the estimation procedure to select the number of degrees of freedom that best fits the data, thus not ruling out leptokurtosis a priori. If the estimate of ν is relatively large, then the Normal may not be a bad approximation.

Because the Sudent-t distribution, with ν finite, may imply no finite unconditional moments for the error process, Nelson(1991) proposes the use of the Generalized Error Distribution (GED). Its density function, for a random variable normalized to have zero mean and unit variance, is:

$$f(z) = \frac{\nu \exp\left[-\frac{1}{2}\left|z/\lambda\right|^{\nu}\right]}{\lambda\, 2^{(1+1/\nu)}\,\Gamma[1/\nu]}, \tag{23}$$

where $\lambda = \left[2^{(-2/\nu)}\,\Gamma[1/\nu]/\Gamma[3/\nu]\right]^{1/2}$. As in the Student-t density, ν controls the tickness of the distribution tail. For $\nu = 2$, the density (23) collapses to the Standard Normal, a result that can be used for a Normality test. For $\nu < 2$, it has ticker tails than the Normal and vice-versa for $\nu > 2$. For $\nu = \infty$, z is uniformly distributed on the interval $\left[-3^{1/2}, 3^{1/2}\right]$.

With correct specification for the functional form of $f(w_t|w_{t-1},\cdots,w_{-k};\theta)$, the log-likelihood function can be written as:

$$\begin{aligned}
\log L(\theta;\cdot) &= \sum_{t=1}^{T} \log\left(f(w_t|w_{t-1},\cdots,w_{-k};\theta)\right) \\
&= \sum_{t=1}^{T} l_t(\cdot), \qquad (24)
\end{aligned}$$

where $l_t(\cdot)$ is the individual likelihood contribution. The function (24) is usually maximized by numerical methods subject to non-negativity constraints whenever necessary. If θ_0 is the true value of the parameters in (24), where $\theta_0 \in \text{int}\,\Theta$, and Θ is a compact subspace of a Euclidean space such that the error process has finite second moments, then, under fairly general conditions (e.g., Weiss(1986)), the maximum likelihood estimate of θ, $\hat{\theta}_T$, converges in distribution as follows:

$$\sqrt{T}\left(\hat{\theta}_T - \theta_0\right) \xrightarrow{d} N\left(0, I_{\theta}^{-1}\right), \tag{25}$$

where $I_{\theta}^{-1} = \left[-E\left(\left.\frac{\partial^2 l_t(\cdot)}{\partial\theta\,\partial\theta'}\right|\Omega_{t-1}\right)\right]^{-1}$ is the inverse of Fisher's Information Matrix. Under a correctly specified model, the latter can be consistently estimated by $\left(T^{-1}\sum_{t=1}^{T}\frac{\partial l_t(\cdot)}{\partial\theta}\frac{\partial l_t(\cdot)}{\partial\theta'}\right)^{-1}$, evaluated at $\hat{\theta}_T$, allowing inference on θ to be conducted using (25).

If there is doubt about the parametric density $f(w_t|w_{t-1},\cdots,w_{-k};\theta)$, but the researcher uses the Normal density, $\hat{\theta}_T$ can still be regarded as the quasi-maximum likelihood estimate of θ. In this case, inference can be conducted using the appropriate correction proposed by Bollerslev and Wooldridge(1992), which relies on:

$$\sqrt{T}\left(\hat{\theta}_T-\theta_0\right)\xrightarrow{d} N\left(0,D^{-1}SD^{-1}\right); \tag{26}$$

where $S=\operatorname{plim}\left(T^{-1}\sum_{t=1}^{T}\frac{\partial l_t(\cdot)}{\partial\theta}\frac{\partial l_t(\cdot)}{\partial\theta'}\right)$, and $D=\operatorname{plim}\left(T^{-1}\sum_{t=1}^{T}-E\left[\frac{\partial^2 l_t(\cdot)}{\partial\theta\,\partial\theta'}\bigg|\,\Omega_{t-1}\right]\right)$, for which consistent estimates can also be constructed making inference feasible.

Testing for ARCH can be easily performed via a Lagrange Multiplier type test proposed by Engle(1982) using the following steps[6]:

1. Run an ordinary least-squares regression to get the residuals $\hat{\varepsilon}_t$. Square them to get $\hat{\varepsilon}_t^2$.

2. Regress $\hat{\varepsilon}_t^2$ on a constant and m of its own lags, obtaining R_u^2 - the uncentered R-squared statistic of this last regression.

3. Under the null that $\varepsilon_t \sim i.i.d.N(0,\sigma^2)$, $T\cdot R_u^2 \xrightarrow{d} \chi_m^2$. Thus, by comparing $T\cdot R_u^2$ with the appropriate entry of a χ^2 distribution table, one can test the null of no ARCH.

3 Volatility in Finance

Volatility is the generic name for conditional standard deviation. In Finance, the term is usually employed to denote the conditional standard deviation of asset returns. Although this subject has evolved considerably in the last twenty years, the market-volatility measures that have been employed in practice are quite naive in statistical sense, since the heteroskedasticity

[6] Other testing procedures are discussed in Bollerslev(1986) and in the survey of Bollerslev, Chou and Kroner(1992).

present in market returns is not usually recognized. Denoting by r_t the demeaned return of a given asset, it is not uncommon for traders to use the following statistic to measure asset-return volatility:

$$\begin{aligned} V_t &= \left(N^{-1} \sum_{i=0}^{N-1} r_{t-i}^2 \right)^{1/2}; \\ t &= N, N+1, \cdots \end{aligned} \tag{27}$$

Notice that (i) V_t^2 is the maximum likelihood estimate of the variance of the return only if r_t is homocedastic and normally distributed, and (ii) V_t is calculated using a fixed window of N observations. Using a window with width N generates an unpleasant property for V_t: there is a (an almost) discrete jump for it when an extreme r_t observation is either included or excluded from the average in (27).

A second commonly used device is the exponential smoothing S_t for the squared asset return:

$$\begin{aligned} S_t &= (1-\lambda) \sum_{j=0}^{\infty} \lambda^j r_{t-j}^2 \\ &= \lambda S_{t-1} + (1-\lambda) r_t^2; \end{aligned} \tag{28}$$

where λ is the decay parameter used to smooth-out lagged squared returns, and it is assumed that $0 < \lambda < 1$. This procedure is identical to the use of a convex combination of lagged S_t and current r_t^2, as shown in the last line of (28). Using $\sqrt{S_t}$ as a volatility measure makes the impact of outliers on $\sqrt{S_t}$ to decrease as time passes, thus it looks smoother than V_t.

A third commonly employed volatility measure is the "implicit volatility" derived from solving the Black and Scholes(1973) formula. A usual problem for this measure is the implicit assumption of log-Normality for the asset price, despite overwhelming empirical evidence to the contrary.

All volatility measures described above disregard the heteroskedasticity present in asset returns[7] and the autocorrelation structure of squared returns. Put in simple terms, they may throw out important information about current an future volatility. Here we go back full circle to Engle's(1982) original idea on Rational Expectations: if there is information on the conditional variance of returns, why not use it?

[7] It could be also argued that they are not based on a proper statistical model.

Models of the ARCH class recognize from the outset that heteroskedasticity is an empirical regularity of financial data. They incorporate this feature of the data by using a simple and ingenious time-series model[8]. In an interesting study, Noh, Engle and Kane(1994) suggest the use of $\sigma_t = \sqrt{E\left[\varepsilon_t^2 \mid \Omega_{t-1}\right]}$ (where ε_t is the innovation in the return series) as a measure of volatility for the return of the S&P500 index. Indeed, these authors find that profits using GARCH$(1;1)$-volatility forecasts significantly exceed transaction costs for near-the-money straddles. This shows that although ARCH has a second order effect on forecasting returns, money could be made by using such information.

GARCH-volatility forecasts have several interesting features. For a stationary GARCH$(1;1)$ model, (13) and (12) above show respectively that variance forecasts (i) have mean reversion to the unconditional variance, and (ii) use the most recent information with the appropriate weights to forecast the variance into the future. Regarding the latter there is a similarity between GARCH-volatility forecasts and the use of the exponential smoothing device. Starting with (9), and assuming that $\alpha; \beta > 0$ and $0 < \alpha + \beta < 1$, we can solve for σ_t^2 in terms of the lagged squared errors to get:

$$\sigma_t^2 = \frac{\omega}{1-\beta} + \alpha \sum_{i=1}^{\infty} \beta^{i-1} \varepsilon_{t-i}^2; \tag{29}$$

which shows that the GARCH model is a way of exponentially smoothing past return innovations; see (28) above. Indeed, if (demeaned) returns are unforecastable, (29) will be a weighted average of lagged squared returns with exponentially decreasing weights.

4 Estimating and Forecasting Volatility in Brazilian Finance Series

4.1 The Data

The data used covers a wide spectrum of Brazilian Finance series: a spot stock-price index - IBOVESPA, of the São Paulo Stock Exchange, the spot

[8] On other grounds, ARCH models were criticized because they were not derived from first principles using Finance theory. Nelson(1990) showed, however, that an ARCH process can be thought as an approximation to a diffusion process, bridging the gap between Finance theory and Econometric models.

price of a popular Brazilian stock - TELEBRAS, traded in the São Paulo Stock Exchange, a spot currency-exchange rate - R$/US$, and a spot popular commodity price - COCOA, with data extracted from the ICCO database. All asset-prices are US$ denominated and data frequency is daily (except for weekends). For the COCOA series the sample covers the period from Jan. 5th, 1990 through Jul. 1st, 1998. For the R$/US$ and TELEBRAS series the sample covers the period from Jul. 4th, 1994 through Jul. 1st, 1998. For the CBOND series the sample period goes from Jul. 18th, 1994 through Jul. 1st, 1998.

There are some missing values for all series. Most are due to holidays. To keep the data frequency uniform across the sample (excluding weekends), missing observations were completed using the most recent quote for each series. Using the transformed data set, the percentage instantaneous return for each series was calculated using a log-difference transformation, i.e., for each price series z_t, $100 \times \Delta \ln(z_t)$ was computed. Plots of the data are presented in Figure 1. As is typical of financial series, they all show sign of heteroskedasticity and volatility clustering.

The return of the US$ shows two distinct patterns of variation during the sample period, reflecting the change in regime from the wide target zones of the beginning of the Real plan to the narrow target zones implemented after the floating of the Mexican Peso. With the exception of the COCOA series, which is traded abroad, all series traded in Brazil vary wildly in relative terms during the Mexican and the Asian crises.

4.2 Estimation Results

Autocorrelation and partial autocorrelation functions for all returns and their squares, as well as other basic statistics for the data, are presented in Table 1. Autocorrelation coefficients are rather small, reflecting no obvious predictability. Indeed, if returns were predictable, there would be arbitrage opportunities for the average investor. Since anyone can be an average investor, there can be no autocorrelation for the return[9]. The highest estimate for the autocorrelation coefficient is 0.12. It happens for the return on the US$ at lag five. Compared to the benchmark of $2/\sqrt{T} = 2/\sqrt{1039} \simeq 0.06$,

[9] Campbell, Lo and McKinley(1994, pp. 85-90) considered spurious autocorrelation arising due to lack of continuous trade. For a sampling interval of one day it is very unlikely for this effect to be relevant.

it is significant. However, the use of this benchmark is only valid under homocedasticity. Indeed, the benchmark underestimates the approximate 95% confidence interval if the data is heteroskedastic, which is probably the case[10].

The autocorrelation coefficients for squared returns show a completely different pattern, reflecting the fact that the conditional variance of returns is predictable. This is corroborated by the fact that all ARCH tests performed reject homocedasticity of returns with great confidence. Also, the Jarque and Bera(1987) normality test rejects Normal returns for all series. The latter has little to do with returns having a skewed distribution, being basically a consequence of outliers in these series; see the very high Kurtosis coefficient for all of them in Table 1a.

The next step was to model the conditional returns taking into account the fact that they are heteroskedastic. The latter is done by fitting the data to a wide variety of popular models of the ARCH class. From the results of this exercise, some stylized facts will surface, being an important component of a later modelling effort. Because of the evidence of non-Gaussian returns, the covariance-matrix correction proposed by Bollerslev and Wooldridge(1992) is employed when the Normal density is used in estimation. The conditional mean of all returns included only a constant term, since returns show no obvious autocorrelation structure. This choice seems appropriate, since, when testing the significance of autoregressive and/or moving-average terms for ARCH-model estimates, the null of a zero coefficient is accepted in all cases[11].

Estimation results for the return on the US$ are presented in Table 2. We first ran a GARCH$(1;1)$ model assuming Gaussian errors. The unit-root test ($\alpha + \beta = 1$ in this case) does not reject the null at usual significance levels (p-value of 0.35). With the exception of the Gaussian TARCH model, the same is observed for all other models used. The EGARCH$(1;1)$ model with Gaussian errors shows no sign of asymmetry of shocks, since the coefficient of $\varepsilon_{t-1}/\sigma_{t-1}$ is not significant. When the GED is used instead of the Normal, we find some evidence of asymmetry, but it is against the leverage effect.

[10] To check if the (instantaneous) returns of these four financial series have first order serial correlation, we regressed them on their first lag; see Table 1a. The reported standard errors, estimated using the procedure in Newey and West(1987), are robust to Heteroskedasticity and serial correlation in the error process. With the exception of the exchange rate series, we find no evidence of first order serial correlation in the return.

[11] For the exchange rate, there is a contradiction between this result and the result in Table 1a, where the Newey and West(1987) robust standard error was used.

Similar, but weaker evidence, is also found for the Gaussian TARCH model. Compared to the Gaussian case, there is an improvement in the AIC and BIC criteria when we use the Student-t or the GED distributions for the error in the GARCH(1;1) and the EGARCH(1;1) model respectively. The estimated degrees of freedom are relatively low (2.64 and 0.98 respectively), a clear sign of leptokurtosis[12]. Hence, we have evidence of fat tails for the conditional distribution, possibly a unit root for the conditional variance, and weak evidence of asymmetry.

Because of the suspicion of a change in target-zone regimes noted earlier, we should be cautious about the evidence of asymmetry and of a unit root for the conditional variance. Indeed, Hamilton and Susmel(1994) point out that a structural break in the variance series may induce a unit root for it. Ignoring changes in regimes may also induce the spurious impression that volatility is related to returns. This may happen because returns may di¤er for the two target-zone regimes[13].

Estimation results for the return of the CBOND are presented in Table 3. We …nd evidence of asymmetry in the variance for the Gaussian EGARCH(1;1), the GED EGARCH(1;1), and the Gaussian TARCH(1;1). All indicate the presence of the leverage e¤ect. As suspected from earlier tests, the distribution of the returns has fat tails: the estimated degrees of freedom for the Student-t GARCH(1;1) and the GED EGARCH(1;1) are 3.71 and 1.01 respectively. The latter is statistically di¤erent from two in hypothesis testing, corroborating our previous …nding of leptokurtosis. If the asymmetric e¤ect is not taken into account the unit-root test rejects the null, but the opposite happens when it is considered. We conclude that there is evidence of asymmetry, favoring the leverage e¤ect, heavy tails for the conditional distribution, and no clear sign of a unit root for the conditional variance.

For the return on COCOA we …nd no asymmetry at all, some evidence of a unit root for the conditional variance, and fat tails; see results in Table 4. The estimated degrees of freedom for the Student-t and GED distributions are respectively 3.32 and 0.977. The latter rejects Normality with high con…dence in hypothesis testing. It is also interesting to note that, regardless of the distribution used, unit-root evidence is not present when we use the

[12] Using the Generalized Error Distribution allows a Normality test as discussed above. At usual levels, conditional Normality is rejected for the return on the US$.

[13] Indeed, average daily returns on the US$ were -0.07% during the wide target-zone regime, increasing to 0.03% during narrow target-zone regime.

EGARCH$(1,1)$ model.

For the return on TELEBRAS we definitely find asymmetry, favoring the leverage effect; see the results in Table 5. This happens regardless of the model or distribution used. There is also a weak sign of a unit root, again rejected whenever asymmetry is considered. Evidence of fat tails is weaker than that for other series, since the estimated degrees of freedom for the Student-t and GED distribution are respectively 6.81 and 1.41. However, formal statistical testing rejected the Normal distribution again when the GED specification is used.

We now turn to SWARCH-model estimation. First, we entertained the two-regime model, with $s_t = 1$ for the low-variance regime, with g_1 normalized to unity, and with $s_t = 2$ when the variance is g_2 greater than that in regime one. Tables 6 through 9 present estimates using the SWARCH model for all four series. Following Hamilton and Susmel(1994), we considered only the case of Student-t and Normal densities.

For the return of the US$, including a first-order autoregressive term for the return did not improve estimation results at all. For the Gaussian SWARCH $(2,2)$ model, the t-test for the significance of the AR(1) coefficient yields a statistic of 0.67[14], not significant at usual levels[15]. There is no evidence of asymmetry in the variance either - the t-statistic for the leverage coefficient is virtually zero. Overall, our preferred model is the SWARCH $(2,4)$, whereas using a Student-t density yields a much higher maximized likelihood value than using the Normal, at the cost of just one more degree of freedom. Indeed, the likelihood-ratio test statistic for comparing the Student-t density with the Normal for the SWARCH $(2,4)$ model is $-2 \times (904.39 - 968.72) = 128.66$, overwhelmingly significant at any reasonable level. Finally, the preferred model for the return on the US$ has a variance in regime two 132.01 times higher than that in the low-variance regime (regime 1).

[14] When the Student-t SWARCH $(2,2)$ or the Gaussian SWARCH $(2,4)$ models were considered, there were problems fitting a first-order autoregression for the mean return, since in all attempts the value of the likelihood could not beat that of the constant return specification; see Table 6. Even when the estimation algorithm converged, the t-test was insignificant for the lagged return. We take these result as evidence that, when the two regimes are considered, there is no first-order autocorrelation for the US$ return.

[15] Consistent with the suggestion in Perron(1989) and Hamilton and Susmel(1994) of an upward bias in the AR(1) coefficient, our previous coefficient estimate was 0.13, while the two-regime estimate is only 0.01 for the Gaussian SWARCH $(2,2)$.

When the Student-t density was used, the estimated transition probability matrix was most of the time in the boundary of its constraint. The same is reported for some estimates in Hamilton and Susmel(1994, footnote 5). To avoid meaningless probability estimates, the model was reparameterized to ensure that $0 \leq p_{ij} \leq 1$, and $\sum_{j=1}^{k} p_{ij} = 1$. Since calculating the respective p_{ij}-estimate standard errors is time consuming, we refrain from doing it here.

For the return on the CBOND our preferred models is the $SWARCH-L(2;3)$; notice that the coefficient of the fourth lagged squared error is not significant. Conforming to our previous evidence, the leverage effect for the CBOND is confirmed using a t-ratio test: 3.45 and 3.33 when using the Student-t and Normal densities respectively. Also based on a t-test, including a first-order autoregressive term for the return is insignificant, with a t-ratio of 1. When testing which density to use, the likelihood-ratio test statistic for comparing the Student-t density with the Normal for the $SWARCH-L(2;3)$ model is $-2 \times [-1944.22 - (-1943.74)] = 0.96$, which is not significant at usual levels. Thus, it makes little difference which one is chosen here. Given our estimate of g_2, CBOND volatility in regime 2 is about 2.2 times higher than that of regime 1; notice that g_2 is statistically different than one at usual levels.

For the return on COCOA our preferred models were the Gaussian $SWARCH(2;4)$ and the Student-t $SWARCH(2;1)$[16]. Confirming our previous results, there is no asymmetry in the variance; a t-ratio smaller than 1 for the leverage coefficient. As it happened for the return of the US$, the estimated transition probability matrix was most of the time in the boundary of its constraints when the Student-t density was used, which lead to the estimation of a reparameterized model for which we do not report standard errors for probability estimates. Last, when we compared the Student-t with the Normal for the $SWARCH(2;4)$, the likelihood-ratio test statistic was $-2 \times [-3968.02 - (-3912.25)] = 111.54$, which overwhelmingly significant at usual levels. Therefore it seems more appropriate to use the Student-t distribution[17]. Given our estimate of g_2 for the Student-t $SWARCH(2;1)$,

[16] Using a Student-t density proved to be difficult in estimation, since we had convergence problems in several occasions. In particular, we could not fit a $SWARCH(2;2)$ model, a candidate for a parsimonious alternative to the $SWARCH(2;4)$. Although there were no convergence problems for the Student-t $SWARCH(2;1)$ (see Table 8), it would be interesting to compare it with the $SWARCH(2;2)$.

[17] A major difference in Student-t- and Gaussian-density estimates is for the transition probability parameters. In the Gaussian case, $\hat{p}_{11} = 0.84$ and $\hat{p}_{22} = 0.24$, whereas for the

COCOA volatility in regime 2 is about 1.7 times higher than that of regime 1; notice that g_2 is statistically different than one at usual levels.

For the return on TELEBRAS our preferred models is the $SWARCH - L(2;4)$. Comparing the Student-t to the Normal yields a likelihood-ratio test statistic of $-2 \times [-2484.14 - (-2475.90)] = 16.48$, which is significant at usual levels. The key difference between Student-t- and Gaussian-density estimates is in the transition probability-matrix parameters. For the Gaussian case, $\hat{p}_{11} = 0.44$ and $\hat{p}_{22} = 0.93$, whereas for the Student-t $\hat{p}_{11} = 0.99$ and $\hat{p}_{22} = 0.99$. The scale parameter and the volatility constant are also different for the two specifications. For the Student-t $\hat{g}_2 = 3.50$ and $\hat{a}_0 = 1.97$, while for the Gaussian case $\hat{g}_2 = 22.75$ and $\hat{a}_0 = 0.10$. Given our estimate of g_2 for the Student-t $SWARCH - L(2;4)$, TELEBRAS volatility in regime 2 is about 1.9 times higher than that of regime 1; notice that g_2 is statistically different than one at usual levels. Conforming to our previous evidence, the leverage effect is present and significant for all estimated models. For our preferred model, the leverage coefficient is 0.30, with a t-ratio of 3.33.

A last modelling effort is made using SWARCH models, in which a three-regime model is entertained. Following Hamilton and Susmel the goal is to allow for an extra regime to capture extreme outliers in the data set; see their discussion in pp. 327-330. This may be useful for Brazilian data since the chances of observing outliers here are much higher than those in developed economies. Estimates using the Gaussian density are reported for the returns of the US$, the CBOND, and TELEBRAS in Table 10. Due to convergence problems, neither returns on COCOA could be estimated using a Gaussian specification, nor could be returns on any of the assets using the Student-t density.

Following our previous results, the returns on the CBOND and TELEBRAS allow for the leverage effect. To make numerical estimation feasible, The order of the ARCH models had to be limited to two[18], which resulted in a $SWARCH - L(3;2)$ model for the returns on the CBOND and TELEBRAS, and a $SWARCH(3;2)$ model for the return on COCOA. The three-regime models were tested against their two-regime counterparts using the likelihood-ratio test statistic[19]. The latter is $-2 \times [883.07 - (909.67)] = 53.20$,

Student-t $\hat{p}_{11} = 0.996$ and $\hat{p}_{22} = 0.994$.

[18] The GAUSS code uses the OPTMUM library. Due to memory restrictions of our current version of GAUSS (3.2), even when the memory extension command was present, it was infeasible to allow for ARCH models with order higher than two.

[19] The overwhelming departure from Normality documented before for all series, raises

$-2\times[-1953.42-(-1946.38)] = 14.08$, and $-2\times[-2485.63-(-2482.63)] = 6.00$, for their return on the US$, on the CBOND and on TELEBRAS respectively. Under correct specification, these test statistics are asymptotically distributed chi-squared with 3, 2, and 1 degrees of freedom respectively, rejecting the two-regime models are rejected at 5% significance.

4.3 Comparing Different ARCH-Model Estimates

This Section focus on comparing goodness-of-fit and forecasting accuracy for ARCH-models. The goodness-of-fit statistics used here are all likelihood based. In particular, we consider the maximum of the log-likelihood function, and the Akaike(1973) and Schwarz(1978) information criteria, which are a function of the former. Comparing forecasting accuracy of different models that predict the conditional mean of a given variable is a simple task. Suppose that we have the sequence $\{y_t; x_t\}_{t=1}^{T+N}$ of realizations of random variables, where y_t is the realization of the explained variable in a regression, and x_t is a vector containing realizations of possible explanatory variables for y_t. In principle, we could consider M different models, indexed by $i = 1, \cdots M$, that hold for the population counterparts of $y_t; x_t$ with error ε_t^i:

$$Y_t = f_i(X_t; \beta_i) + \varepsilon_t^i; \tag{30}$$

where $t = 1, \cdots T$. Based on some optimality criteria, these M models could be estimated, resulting in $\hat{\beta}_i$ - model i's estimate for the conditional-mean parameter β_i. Conditional on x_t, for $t = T+1, \cdots N$, the out-of-sample forecasting accuracy of these M models could be compared using some loss-function. In particular, if the mean-squared-error function is considered, the following statistic for all M models could be calculated:

$$\begin{aligned} MSE^i &= N^{-1} \sum_{t=T+1} \left(y_t - f_i\left(x_t; \hat{\beta}_i \right) \right)^2; \\ i &= 1, \cdots M. \end{aligned} \tag{31}$$

Under the usual caveats, the "best" model would be the one with the smallest value for (31).

the issue that the likelihood-ratio test is not appropriate in this case. Still, we use it here as a benchmark.

Unfortunately, this same procedure cannot be replicated if the goal is to measure forecasting accuracy for the conditional variance (the same applies to volatility). This happens because the conditional variance is not a random variable for which we can collect realizations to form statistics such as (31). On the contrary, σ_t^2 is an unknown time-varying parameter that could, at best, be estimated consistently when the true Data Generating Process (DGP) is known. In general, since the DGP is unknown, there is no hope of even getting a consistent estimate.

Recognizing this problem, different authors have proposed tracking down not the conditional variance but some other variable, which may be the same for all volatility forecasts. For example, Heynen and Kat(1994) use what they label "realized volatility," a degrees-of-freedom corrected version of (27). Others have proposed using implicit volatility; see Engle and Mustafa(1992). On the other hand, using the definition of conditional variance, i.e., $\sigma_t^2 = E[\varepsilon_t^2 | \Omega_{t-1}]$, Hamilton and Susmel propose comparing each model's variance forecast with what it is supposed to track down. Thinking in terms of one-step-ahead forecast errors, they propose comparing σ_t^2 with ε_t^2, using their respective estimates. We follow Hamilton and Susmel in assessing the forecast accuracy of our volatility estimates by using four different loss functions: mean-squared-error (MSE), mean-absolute-error (MAE), mean-squared-log-error $[LE]^2$, and mean-absolute-log-error $|LE|$. Results are presented in Tables 11 through 14.

For the return of the US\$, the $EGARCH(1,1)$ performs very well. For the $[LE]^2$ and $|LE|$ loss functions, the best model is the Gaussian $EGARCH(1,1)$. When the MAE is used the GED $EGARCH(1,1)$ is the best, followed closely by the Gaussian $SWARCH(3,2)$, which is the best when the MSE is used.

For the return on the CBOND, the $GARCH(1,1)$ using either a Gaussian or Student-t specification performs best for the $[LE]^2$ and $|LE|$ functions. For the MAE or the MSE functions the Gaussian $EGARCH(1,1)$ is the best model. It is worth mentioning that the Gaussian $SWARCH(3,2)$ does also well when we used the MSE, the MAE, and the $|LE|$ functions.

For the return on COCOA, the Gaussian $EGARCH(1,1)$ performs very well, with the smallest loss-function value when we used the MSE, the $[LE]^2$, and the $|LE|$ functions. For the MAE function, the best model is also the $EGARCH(1,1)$, when the Generalized Error Distribution is used. In this case, the Gaussian $EGARCH(1,1)$ gets the second smallest statistic. It is worth mentioning that the Gaussian $GARCH(1,1)$ does also well when we used the $[LE]^2$ and the $|LE|$ functions.

For the return on TELEBRAS, the Gaussian GARCH $(1;1)$ performs best for the $[LE]^2$ and $|LE|$ functions. For the MAE function, the Gaussian EGARCH $(1;1)$ is the best model, but for the MSE function, the Gaussian TARCH $(1;1)$ is the best model.

Overall, the Gaussian EGARCH $(1;1)$ performed very well. Similar results are obtained by Pagan and Schwert(1990) and Engle and Ng(1991). Since for the EGARCH model the impact of squared errors on the conditional variance is exponential, it is thought to react too much to lagged standardized errors. This may be bad if large errors are infrequent, with the model over-predicting the variance in response to a sequence of small errors. However, if large errors are common, this feature may not be bad, since it also matters how well the model forecast outliers.

The forecasting performance SWARCH models was not encouraging compared to other models of the ARCH class. Hamilton and Susmel criticize standard GARCH models for overestimating the persistence of volatility[20]. Indeed, they write in p. 316 that "Engle and Mustafa(1992) concluded on the basis of stock option prices that the volatility consequences of the 1987 crash disappeared more rapidly than is suggested by the ... [behavior of the Student-t TARCH $(1;1)$ model]. Lamoureux and Lastrapes(1993) presented related evidence based on earlier data that standard GARCH models overforecast the persistence in volatility." If this is true, our forecasting results show that overforecasting actually helped ARCH models that neglect regime switching. This may be related to the frequency and size of outliers for Brazilian data. If outliers are rare, it is probably not very good to have a model which frequently overestimates the volatility of regular standardized errors. However, if outliers are frequent, overforecasting will hurt the forecast of mid-sized errors but probably benefit the forecast of outliers. Since these make a large contribution to the average forecasting error, the net result may be favorable to models with this feature. Despite this conjecture, It is worth noting that for the return on the US$, where there are clearly distinct regimes, the Gaussian SWARCH$(3;2)$ performed well, as expected.

Finally, we present goodness-of-fit statistics for most regressions in Tables 15 through 18. For the return on the US$ and on TELEBRAS, the best

[20] For the value-weighted portfolio of the NYSE, Hamilton and Susmel find the persistence coefficient for several GARCH models to be in $[0{:}96; 0{:}99]$. For SWARCH models the interval was $[0{:}42; 0{:}59]$; see the results in the last column of their Table 1.

model is the Student-t SWARCH$(2;4)$, although for TELEBRAS, using the Schwarz criterium, one would have chosen the Student-t GARCH$(1;1)$, because the former has too many parameters. For the return on the CBOND the best model is the Student-t GARCH$(1;1)$, and for the return on the COCOA the best is the GED EGARCH$(1;1)$. These results partially rehabilitate models in the SWARCH class, although it deserves further investigation why their forecasting performance is not as good as their goodness-of-fit statistics.

5 Conclusions and Further Research

The goal of this paper was to present a comprehensive empirical analysis of the return and conditional variance of four Brazilian financial series using models of the ARCH class. To discuss the empirical results in greater depth, a self-contained theoretical Section presents ARCH models in a way that it is useful for the applied researcher. References to complete surveys are also given.

The empirical results show a distinct behavior for these four financial series. Although all series share ARCH and are leptokurtic relative to the Normal, the return on the US$ has clearly regime switching and no asymmetry for the variance, the return on COCOA has no asymmetry, while the returns on the CBOND and TELEBRAS have clear signs of asymmetry favoring the leverage effect. All these stylized facts were modelled using the ARCH class. Regarding forecasting, the best model overall was the EGARCH$(1;1)$, in its Gaussian version. Different versions of the GARCH$(1;1)$ also performed well, while the SWARCH only did well for the return on the US$, which has a distinct pattern of regimes for the sample period. Regarding goodness-of-fit statistics, the SWARCH model did well, followed closely by the Student-t GARCH$(1;1)$.

Understanding forecasting results deserves further investigation. Maybe a comparison between models of the EGARCH and the SWARCH family would be useful, since it may shed light in why the former does so well and the latter does not. This is particularly intriguing in light of our goodness-of-fit results.

References

[1] Akaike, H. (1973) Information Theory and an Extension of the Maximum Likelihood Principle, in: B.N. Petrov and F. Csáki, eds., Second International Symposium on Information Theory. Akadémiai Kiadó: Budapest.

[2] Black, F. (1976) Studies of Stock Price Volatility Changes, Proceedings from the American Statistical Association, Business and Economic Statistics Section, 177-181.

[3] Black, F. and M.Scholes (1973) The Pricing of Options and Corporate Liabilities, Journal of Political Economy, 81, 637-659.

[4] Bollerslev, Tim (1986). Generalized Autoregressive Conditional Heteroskedasticity, Journal of Econometrics 31, 307-327.

[5] Bollerslev, T. (1987) A Conditional Heteroskedastic Time Series Model for Speculative Prices and Rates of Return, Review of Economics and Statistics, 69, 542-547.

[6] Bollerslev Tim, Ray Y. Chou, and Kenneth F. Kroner (1992) ARCH Modeling in Finance: A Review of the Theory and Empirical Evidence, Journal of Econometrics 52, 559.

[7] Bollerslev, Tim and Je¤rey M. Wooldridge (1992) Quasi-Maximum Likelihood Estimation and Inference in Dynamic Models with Time Varying Covariances, Econometric Reviews, 11, 143-172.

[8] Bollerslev, Tim, Robert F. Engle and Daniel B. Nelson (1994) ARCH Models, in Chapter 49 of Handbook of Econometrics, Volume 4, North-Holland.

[9] Box, G.E.P., and G.M. Jenkins (1976) Time Series Analysis: Forecasting and Control. Holden day: San Francisco, CA. Second Edition.

[10] Campbell, J.Y., Lo, A.W., and MacKinlay, A.C. (1997), The Econometrics of Financial Markets. Princeton: Princeton University Press.

[11] Diebold, Francis X., 1986. Modeling the persistence of conditional variances: A comment, Econometric Reviews 5, 51-56.

[12] Engle, Robert F. (1982), Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. In‡ation, Econometrica, 50, 987-1008.

[13] Engle, Robert F. (1983), Estimates of the Variance of U.S. In‡ation Based upon the ARCH Model, Journal of Money, Credit and Banking, 15.

[14] Engle, Robert F. (1995), ARCH: Selected Readings. Oxford: Oxford University Press.

[15] Engle, R.F. and G. Gonzalez-Rivera (1991) Semiparametric ARCH Models, Journal of Business and Economic Statistics, 9, 345-359.

[16] Engle, R.F. and C. Mustafa (1992) Implied ARCH Models from Options Prices, Journal of Econometircs, 52, 289-311.

[17] Engle, Robert F. and Victor K. Ng (1993) Measuring and Testing the Impact of News on Volatility, Journal of Finance, 48, 1022-1082.

[18] Friedman, M. (1977), Nobel Lecture: In‡ation and Unemployment, Journal fo Political Economy, 85.

[19] Gallant, A.R. and G.Tauchen (1989) Semi Non-Parametric Estimation of Cinditionally Constrained Heterogeneous Processes: Asset Pricing Applications, Econometrica, 57, 1091-1120.

[20] Gallant, A.R., D.A. Hsieh and G. Tauchen (1991) On Fitting a Recalcitrant Series: The Pound/Dollar Exchange Rate 1974-83, in: W.A. Barnett, J. Powell and G. Tauchen, eds., Nonparametric and Semiparametric Methods in Econometrics and Statistics. Cambridge University Press: Cambridge.

[21] Gallant, A.R., P.E. Rossi and G. Tauchen (1992) Stock Prices and Volume, Review of Financial Studies, 5, 199-242.

[22] Gallant, A.R., P.E. Rossi and G. Tauchen (1993) Non linear Dynamic Structures, Econometrica, 61, 871-907.

[23] Glosten, L.R., R. Jagannathan, and D. Runkle (1993) On the Relation between the Expected Value and the Volatility of the Normal Excess Return on Stocks, Journal of Finance, 48, 1779-1801.

[24] Granger, C.W.J. and Andersen, A. (1978), An Introduction to Bilinear Time-Series Models. Göttingen.

[25] Hamilton, James D. (1994) Time Series Analysis, Princeton University Press.

[26] Hamilton, James D., and Raul Susmel (1994) Autoregressive Conditional Heteroskedasticity and Changes in Regime, Journal of Econometrics, 64, 307-333.

[27] Jarque, C. and Bera, A. (1987), A Test for Normality of Observations and Regression Residuals, International Statistical Review, 55, 163-172.

[28] Lamoureux, Christopher G. and William D. Lastrapes (1990) Persistence in Variance, Structural Change and the GARCH model. Journal of Business and Economic Statistics 8, 225-234.

[29] Lamoureux, Christopher G. and William D. Lastrapes (1993) Forecasting Stock Return Variance: Toward an Understanding of Stochastic Implied Volatilities, Review of Financial Studies, 5, 293-326.

[30] Mandelbrot, B. (1963) The Variation of Certain Speculative Prices, Journal of Business, 36, 394-419.

[31] Nelson, D.B. (1990) ARCH Models as Di¤usion Aproximations, Journal of Econometrics, 45, 7-38.

[32] Nelson, Daniel B. (1991) Conditional Heteroskedasticity in Asset Returns: A New Approach, Econometrica, 59, 347-370.

[33] Newey, Whitney and Kenneth West (1987) A Simple Positive Semi-De…nite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix, Econometrica, 55, 703-708.

[34] Okun, A. (1971), The Mirage of Steady In‡ation, Brookings Papers on Economic Activity, 2.

[35] Pagan, A.R. and G.W. Schwert (1990) Alternative Models for Conditional Stock Volatility, Journal of Econometrics, 45, 267-290.

[36] Perron, Pierre, 1989. The Great Crash, the Oil Price Shock, and the Unit-Root Hypothesis. Econometrica 57, 1361-1401.

[37] Schwarz, G. (1978) Estimating the Dimension of a Model. Annals of Statistics, 6, 461-464.

[38] Tauchen, George (1986). Statistical Properties of Generalized Method-of-Moments Estimators of Structural Parameters Obtained From Financial Market Data, Journal of Business & Economic Statistics, 4, 397-416.

[39] Weiss, A.A. (1986) Asymptotic Theory for ARCH Models: Estimation and Testing, Econometric Theory, 2 107-131.

[40] Zakoian, J.M. (1990) Threshold Heteroskedastic Models, manuscript, CREST, INSEE, Paris.

Table 1: Stylized Facts of Brazilian Returns

a) Descriptive Statistics

	US$	CBOND	COCOA	TELEBRAS
Mean Daily Return (%)	0.023	0.067	0.032	0.114
Skewness	0.181	-0.545	0.405	0.685
Kurtosis	18.641	17.727	9.483	15.741
Jarque-Bera Test (P-value)	0.000	0.000	0.000	0.000
Unconditional standard deviation	0.282	1.875	1.633	3.227
ARCH(5) Test (P-value)	0.000	0.000	0.000	0.000
AR(1) Robust *t*-test (P-value)	0.006	0.928	0.200	0.99
Number of Observations	1039	1094	2210	1039

Notes:
(1) AR(1) coefficient standard error calculated using the procedure in Newey and West (1987).

b) Autocorrelation and Partial Autocorrelation Functions of Returns and Squared Returns

Returns	US$	CBOND	COCOA	TELEBRAS
$r_1 (a_1)$	0.106 (0.106)	-0.007 (-0.007)	-0.039 (-0.039)	-0.002 (-0.002)
$r_2 (a_2)$	-0.012 (-0.023)	0.017 (0.017)	-0.042 (-0.044)	-0.030 (-0.030)
$r_3 (a_3)$	-0.082 (-0.079)	-0.107 (-0.107)	-0.007 (-0.010)	-0.042 (-0.042)
$r_4 (a_4)$	0.037 (0.055)	-0.013 (-0.015)	0.013 (0.011)	-0.025 (-0.026)
$r_5 (a_5)$	0.120 (0.110)	-0.037 (-0.035)	-0.010 (-0.010)	-0.080 (-0.083)
$2/\sqrt{T}$	0.062	0.060	0.043	0.062

Squared Returns	US$	CBOND	COCOA	TELEBRAS
$r_1 (a_1)$	0.234 (0.234)	0.271 (0.271)	0.123 (0.123)	0.145 (0.145)
$r_2 (a_2)$	0.157 (0.108)	0.287 (0.230)	0.042 (0.027)	0.217 (0.200)
$r_3 (a_3)$	0.160 (0.109)	0.234 (0.128)	0.028(0.020)	0.154 (0.106)
$r_4 (a_4)$	0.160 (0.097)	0.060 (-0.087)	0.017 (0.010)	0.084 (0.014)
$r_5 (a_5)$	0.162 (0.091)	0.074 (-0.008)	0.058 (0.055)	0.105 (0.047)
$2/\sqrt{T}$	0.062	0.060	0.043	0.062

Notes:
(1) ρ_i and a_i denote respectively i-th order autocorrelation and partial-autocorrelation coefficient estimates.

Table 2: Basic ARCH Estimates for the Return on the US$

Mean	Constant	0.028 (13.41)	0.031 (9.03)	0.024 (15.73)	0.019 (18.13)	0.029 (14.42)
Variance	Constant	9.96.E-5 (1.51)	-0.32 (-3.98)	0.0002 (2.72)	166.91 (0.02)	8.38E-5 (1.40)
	e_{t-1}^2	0.22 (3.37)		0.28 (3.37)		0.30 (3.89)
	$d_{t-1} \cdot \varepsilon_{t-1}^2$					-0.18 (-1.76)
	s_{t-1}^2	0.80 (16.22)		0.82 (33.26)		0.81 (19.35)
	$\lvert e_{t-1} \rvert / s_{t-1}$		0.39 (4.13)		0.15 (7.81)	
	e_{t-1} / s_{t-1}		0.09 (1.37)		0.43 (2.69)	
	$\ln\left(s_{t-1}^2\right)$		0.99 (126.41)		0.999 (443.99)	
Estimation Method		ML	ML	ML	ML	ML
Model		GARCH(1,1)	EGARCH(1,1)	GARCH (1,1)	EGARCH (1,1)	TARCH(1,1)
Distribution		$N(\cdot)$	$N(\cdot)$	$t(\cdot)$	GED	$N(\cdot)$
Log L		872.31	876.79	945.66	931.75	882.56
AIC		-1.6714	-1.6781	-1.8107	-1.782	-1.689
BIC		-1.6524	-1.6543	-1.7869	-1.753	-1.665
Sample		8/7/94 1/7/98	8/7/94 1/7/98	8/7/94 1/7/98	8/7/94 1/7/94	8/7/94 1.7/98
Unit-root test (P-value)		0.35	0.45	0.17	0.98	0.047
Degrees of freedom				2.64 (10.82)	0.82 (22.25)	

Notes:
(1) *t*-statistics in parentheses.
(2) Gaussian EGARCH(1,1) estimates equation (15), while EGARCH(1,1) with the GED specification estimates equation (16).
(3) Gaussian GARCH(1,1), TARCH(1,1) and EGARCH(1,1) estimates use the Bollerslev and Wooldridge(1992) quasi-maximum likelihood asymptotic variance-covariance matrix.

Table 3: Basic ARCH Estimates for the Return on the CBOND

Mean	Constant	0.134 (3.72)	0.08 (2.09)	0.15 (4.51)	0.098 (3.31)	0.08 (2.05)
Variance	Constant	0.092 (2.15)	-0.15 (-3.08)	0.062 (2.94)	1.20 (5.12)	0.139 (3.05)
	e_{t-1}^{2}	0.18 (1.66)		0.11 (4.70)		0.03 (1.42)
	$d_{t-1} \cdot \varepsilon_{t-1}^{2}$					0.23 (2.02)
	s_{t-1}^{2}	0.81 (10.08)		0.88 (41.92)		0.80 (14.32)
	$\left\|e_{t-1}\right\| / s_{t-1}$		0.27 (3.01)		0.17 (4.98)	
	e_{t-1} / s_{t-1}		-0.15 (-2.55)		-0.47 (-2.82)	
	$\ln\left(s_{t-1}^{2}\right)$		0.93 (42.84)		0.96 (91.42)	
Estimation Method		ML	ML	ML	ML	ML
Model		GARCH(1,1)	EGARCH(1,1)	GARCH(1,1)	EGARCH(1,1)	TARCH(1,1)
Distribution		$N(\cdot)$	$N(\cdot)$	$t(\cdot)$	GED	$N(\cdot)$
Log L		-2001.363	-1981.85	-1928.77	-1937.33	-1980.54
AIC		3.6661	3.6323	3.5352	3.553	3.630
BIC		3.6844	3.6551	3.5439	3.580	3.653
Sample		22/4/94 1/7/98	22/4/94 1/7/98	22/4/94 1/7/98	22/4/94 1/7/98	8/7/94 1/7/98
Unit-root test (P-value)		0.68	0.005	0.63	0.0008	0.0006
Degrees of freedom				3.71 (8.47)	1.01 (24.06)	

Notes:
(1) *t*-statistics in parentheses.
(2) Gaussian EGARCH(1,1) estimates equation (15), while EGARCH(1,1) with the GED specification estimates equation (16).
(3) Gaussian GARCH(1,1), TARCH(1,1) and EGARCH(1,1) estimates use the Bollerslev and Wooldridge(1992) quasi-maximum likelihood asymptotic variance-covariance matrix.

Table 4: Basic ARCH Estimates for the Return on COCOA

Mean	Constant	0.029 (0.94)	0.065 (1.94)	-0.014 (-0.57)	0.000 (0.000)	0.037 (1.19)
Variance	Constant	0.017 (2.03)	-0.050 (-311)	0.036 (3.35)	-5.68 (-1.70)	0.016 (2.03)
	e_{t-1}^{2}	0.03 (3.99)		0.04 (4.55)		0.04 (2.03)
	$\mathrm{d}_{t-1} \cdot \varepsilon_{t-1}^{2}$					-0.01 (-0.80)
	s_{t-1}^{2}	0.96 (119.42)		0.95 (115.10)		0.96 (128.05)
	$\lvert e_{t-1} \rvert / s_{t-1}$		0.081 (3.54)		0.044 (4.64)	
	e_{t-1} / s_{t-1}		0.010 (0.54)		0.21 (0.96)	
	$\ln\left(s_{t-1}^{2}\right)$		0.993 (247.04)		0.995 (364.97)	
Estimation Method		ML	ML	ML	ML	ML
Model		GARCH(1,1)	EGARCH(1,1)	GARCH (1,1)	EGARCH(1,1)	TARCH(1,1)
Distribution		$N(\cdot)$	$N(\cdot)$	$t(\cdot)$	GED	$N(\cdot)$
Log L		-4078.65	-4090.70	-3910.23	-3889.42	-4077.27
AIC		3.6947	3.7065	3.5432	3.519	3.694
BIC		3.7050	3.7194	3.5561	3.534	3.707
Sample		11/10/90 1/7/98	11/10/90 1/7/98	11/10/90 1/7/98	5/1/90 1/7/98	8/7/94 1/7/98
Unit-root test (P-value)		0.20	0.07	0.30	0.08	0.9119
Degrees of freedom				3.32 (11.61)	0.977 (30.13)	

Notes:

(1) *t*-statistics in parentheses.

(2) Gaussian EGARCH(1,1) estimates equation (15), while EGARCH(1,1) with the GED specification estimates equation (16).

(3) Gaussian GARCH(1,1), TARCH(1,1) and EGARCH(1,1) estimates use the Bollerslev and Wooldridge(1992) quasi-maximum likelihood asymptotic variance-covariance matrix.

(4) Due to convergence problems the EGARCH (1,1) model with the GED specification uses a reparameterized intercept for the variance, which includes the term $-a\, E\left|Z_{t-1}\right|$

Table 5: Basic ARCH Estimates for the Return on TELEBRAS

Mean	Constant	0.28 (4.28)	0.18 (2.55)	0.28 (4.02)	0.156 (2.22)	0.177 (2.53)
Variance	Constant	0.35 (3.23)	-0.07 (-1.87)	0.33 (3.24)	2.10 (14.19)	0.52 (3.99)
	e_{t-1}^{2}	0.18 (3.97)		0.19 (5.64)		0.04 (1.45)
	$d_{t-1} \cdot \varepsilon_{t-1}^{2}$					0.22 (3.40)
	s_{t-1}^{2}	0.79 (19.30)		0.79 (24.15)		0.79 (19.85)
	$\left\|e_{t-1}\right\| / s_{t-1}$		0.26 (4.55)		0.24 (6.07)	
	e_{t-1} / s_{t-1}		-0.15 (-3.69)		-0.50 (-2.98)	
	$\ln\left(s_{t-1}^{2}\right)$		0.93 (51.10)		0.93 (5957)	
Estimation Method		ML	ML	ML	ML	ML
Model		GARCH(1,1)	EGARCH(1,1)	GARCH (1,1)	EGARCH(1,1)	TARCH(1,1)
Distribution		$N(\cdot)$	$N(\cdot)$	$t(\cdot)$	GED	$N(\cdot)$
Log L		-2498.84	-2483.38	-2485.67	-2488.63	-2483.45
AIC		4.8178	4.7899	4.7944	4.802	4.79
BIC		4.8368	4.8137	4.8034	4.831	4.81
Sample		8/7/94 1/7/98	8/7/94 1/7/98	8/7/94 1/7/98	8/7/94 1/7/98	8/7/94 1/7/98
Unit-root test (P-value)		0.12	0.000	0.36	0.00003	0.0000
Degrees of freedom				6.81 (4.96)	1.41 (17.52)	

Notes:

(1) *t*-statistics in parentheses.

(2) Gaussian EGARCH(1,1) estimates equation (15), while EGARCH(1,1) with the GED specification estimates equation (16).

(3) Gaussian GARCH(1,1), TARCH(1,1) and EGARCH(1,1) estimates use the Bollerslev and Wooldridge(1992) quasi-maximum likelihood asymptotic variance-covariance matrix.

Table 6: Two-regime SWARCH and SWARCHL Estimates for the Return on the US$

Constant (mean)	0.026 (0.002)	0.026 (0.002)	0.026 (0.002)	0.022 (0.002)	0.022 (0.002)	0.022 (0.002)
AR(1)		0.02 (0.03)				
Const. (variance)	0.001 (0.0002)	0.001 (0.0001)	0.002 (0.0002)	0.003 (0.0007)	0.003 (0.0007)	0.002 (0.0005)
$\tilde{u}^2_{t-1}$	0.27 (0.06)	0.27 (0.06)	0.30 (0.08)	0.39 (0.14)	0.39 (0.14)	0.35 (0.11)
$\tilde{u}^2_{t-2}$	0.19 (0.05)	0.20 (0.05)	0.21 (0.06)	0.26 (0.11)	0.26 (0.11)	0.17 (0.08)
$\tilde{u}^2_{t-3}$	0.08 (0.04)	0.08 (0.04)			0.01 (0.03)	0.000 (0.05)
$\tilde{u}^2_{t-4}$	0.18 (0.05)	0.17 (0.05)				0.15 (0.07)
Degrees of freedom				2.77 (0.30)	2.78 (0.30)	3.02 (0.35)
g_2	98.23 (18.12)	98.84 (18.04)	106.77 (15.60)	130.75 (24.91)	130.36 (25.02)	132.01 (26.80)
$d_{t-1} \cdot \tilde{u}^2_{t-1}$						
p_{11}	0.99 (0.005)	0.989 (*)	0.983 (*)	0.999 (*)	0.999 (*)	0.999 (*)
p_{22}	0.97 (0.011)	0.973 (*)	0.959 (*)	0.998 (*)	0.999 (*)	0.999 (*)
$Erg.\ P_1$ $Erg.\ P_2$	0.71 0.29	0.71 0.29	0.71 0.29	0.64 0.36	0.64 0.36	0.63 0.37
Log L	904.39	897.32	883.07	961.62	961.64	968.72
Distribution	$N(\cdot)$	$N(\cdot)$	$N(\cdot)$	$t(\cdot)$	$t(\cdot)$	$t(\cdot)$

Notes:
(1) Standard Errors in parentheses.
(2) (*) Denotes that the probability-parameters in the model were changed to satisfy boundary constraints. Standard-error estimates are feasible but not attempted.

Table 7: Two-Regime SWARCH and SWARCHL Estimates for the Return on the CBOND

Constant (mean)	0.15 (0.04)	0.14 (0.04)	0.13 (0.04)	0.14 (0.04)	0.14 (0.04)
AR(1)			0.01 (0.03)	0.03 (0.03)	
Const. (variance)	0.43 (0.08)	0.44 (0.08)	0.44 (0.07)	0.43 (0.11)	0.45 (0.11)
$\tilde{u}^2_{t-1}$	0.03 (0.06)	0.05 (0.06)	0.13 (0.07)	0.06 (0.08)	0.07 (0.08)
$\tilde{u}^2_{t-2}$	0.17 (0.06)	0.16 (0.06)	0.28 (0.06)	0.15 (0.06)	0.15 (0.06)
$\tilde{u}^2_{t-3}$	0.19 (0.07)	0.22 (0.06)		0.22 (0.07)	0.21 (0.06)
$\tilde{u}^2_{t-4}$	0.06 (0.04)				
Degrees of freedom				16.22 (26.27)	13.92 (18.01)
g_2	5.64 (0.80)	5.56 (0.77)	6.08 (0.75)	4.97 (1.27)	4.80 (1.26)
$d_{t-1} \cdot \tilde{u}^2_{t-1}$	0.37 (0.11)	0.38 (0.11)	0.33 (0.12)	0.41 (0.12)	0.40 (0.12)
p_{11}	0.78 (0.08)	0.75 (0.10)	0.69 (*)	0.77 (0.24)	0.66 (0.23)
p_{22}	0.37 (0.13)	0.35 (0.12)	0.37 (*)	0.38 (0.15)	0.35 (0.16)
Ergodic P_1	0.74	0.72	0.67	0.65	0.66
Ergodic P_2	0.26	0.28	0.33	0.35	0.34
Log L	-1942.91	-1944.22	-1953.43	-1943.26	-1943.74
Distribution	$N(\cdot)$	$N(\cdot)$	$N(\cdot)$	$t(\cdot)$	$t(\cdot)$

Notes:
(1) Standard Errors in parentheses.
(2) (*) Denotes that the probability-parameters in the model were changed to satisfy boundary constraints. Standard-error estimates are feasible but were not attempted.

Table 8: Two-Regime SWARCH and SWARCHL Estimates for the Return on COCOA

Const. (mean)	-0.001 (0.01)	-0.01 (0.03)	-0.01 (0.03)	-0.01 (0.03)	-0.01 (0.03)	-0.001 (0.02)	-0.007 (0.02)	-0.006 (0.02)
AR(1)					-0.02 (0.02)			
Const. (variance)	0.76 (0.11)	0.76 (0.11)	0.72 (0.09)	0.72 (0.08)	0.73 (0.08)	1.43 (0.18)	1.44 (0.17)	1.49 (0.17)
$\tilde{u}^2_{t-1}$	0.17 (0.05)	0.14 (0.06)	0.14 (0.06)	0.11 (0.05)	0.13 (0.05)	0.08 (0.03)	0.08 (0.03)	0.08 (0.03)
$\tilde{u}^2_{t-2}$	0.05 (0.05)	0.04 (0.04)	0.01 (0.02)	0.02 (0.03)	0.02 (0.04)	0.000 (0.03)	0.000 (0.02)	
$\tilde{u}^2_{t-3}$			0.15 (0.04)	0.12 (0.04)	0.12 (0.04)	0.03 (0.03)	0.03 (0.03)	
$\tilde{u}^2_{t-4}$				0.13 (0.04)	0.12 (0.04)	0.03 (0.04)		
g_2	7.91 (0.69)	7.89 (0.68)	8.04 (0.76)	8.16 (0.87)	8.27 (0.90)	2.86 (0.40)	2.85 (0.32)	2.88 (0.32)
p_{11}	0.76 (*)	0.76 (*)	0.79 (*)	0.84 (*)	0.84 (*)	0.996 (*)	0.996 (*)	0.996 (*)
p_{22}	0.30 (*)	0.31 (*)	0.26 (*)	0.24 (*)	0.24 (*)	0.994 (*)	0.995 (*)	0.994 (*)
$d_{t-1} \cdot \tilde{u}^2_{t-1}$		0.05 (0.09)	0.05 (0.07)	0.04 (0.07)				
Degrees of Freedom						3.55 (0.33)	3.58 (0.33)	3.58 (0.33)
$Erg.\ P_1$	0.75	0.74	0.78	0.83	0.83	0.62	0.61	0.61
$Erg.\ P_2$	0.25	0.26	0.22	0.17	0.17	0.38	0.39	0.39
Log L	-3981.12	-3980.91	-3974.06	-3968.02	-3974.15	-3912.25	-3912.90	-3913.44
Distrib.	$N(\cdot)$	$N(\cdot)$	$N(\cdot)$	$N(\cdot)$	$N(\cdot)$	$t(\cdot)$	$t(\cdot)$	$t(\cdot)$

Notes:
(1) Standard Errors in parentheses.
(2) (*) Denotes that the probability-parameters in the model were changed to satisfy boundary constraints. Standard-error estimates are feasible but were not attempted.

Table 9: Two-Regime SWARCH and SWARCHL Estimates for the Return on TELEBRAS

Constant (mean)	0.28 (0.07)	0.17 (0.06)	0.18 (0.06)	0.25 (0.08)	0.27 (0.07)	0.27 (0.07)	0.27 (0.07)
AR(1)		0.01 (0.03)					
Const. (variance)	2.25 (0.27)	0.10 (0.03)	0.10 (0.03)	0.30 (0.19)	1.97 (0.31)	2.17 (0.31)	2.41 (0.31)
$\tilde{u}^2_{t-1}$	0.000 (0.03)	0.03 (0.1)	0.03 (0.04)	0.02 (0.04)	0.000 (0.02)	0.000 (0.02)	0.001 (0.03)
$\tilde{u}^2_{t-2}$	0.17 (0.04)	0.20 (0.06)	0.20 (0.05)	0.22 (0.06)	0.17 (0.05)	0.16 (0.05)	0.17 (0.05)
$\tilde{u}^2_{t-3}$		0.18 (0.06)	0.18 (0.05)	0.16 (0.05)	0.07 (0.04)	0.07 (0.04)	
$\tilde{u}^2_{t-4}$		0.22 (0.06)	0.22 (0.06)		0.09 (0.05)		
p_{11}	0.98 (0.009)	0.43 (0.10)	0.44 (0.10)	0.40 (0.13)	0.99 (0.01)	0.99 (*)	0.99 (0.007)
p_{22}	0.98 (0.01)	0.93 (0.02)	0.93 (0.02)	0.90 (0.04)	0.99 (0.01)	0.99 (*)	0.99 (0.008)
g_2	4.25 (0.57)	23.41 (7.64)	22.75 (7.25)	13.16 (7.15)	3.50 (0.61)	3.54 (0.58)	3.87 (0.59)
$d_{t-1} \cdot \tilde{u}^2_{t-1}$	0.33 (0.08)	0.37 (0.11)	0.37 (0.10)	0.44 (0.11)	0.30 (0.09)	0.31 (0.09)	0.31 (0.09)
$Erg.\quad P_1$	0.48	0.11	0.11	0.15	0.41	0.42	0.47
$Erg.\quad P_2$	0.52	0.89	0.89	0.85	0.58	0.58	0.53
Log L	-2485.63	-2490.41	-2484.14	-2500.07	-2475.90	-2478.36	-2481.17
Degrees of Freedom					10.01 (3.11)	10.93 (3.81)	11.33 (4.06)
Distribution	$N(\cdot)$	$N(\cdot)$	$N(\cdot)$	$N(\cdot)$	$t(\cdot)$	$t(\cdot)$	$t(\cdot)$

Notes:
(1) Standard Errors in parentheses.
(2) (*) Denotes that the probability-parameters in the model were changed to satisfy boundary constraints. Standard-error estimates are feasible but were not attempted.

Table 10: Three-Regime SWARCH and SWARCHL Estimates with Gaussian Errors

	R$	CBond	TELEBRAS
Constant (mean)	0.026 (0.002)	0.13 (0.04)	0.27 (0.07)
AR(1)		0.01 (0.03)	
Const. (variance)	0.001 (0.0001)	0.56 (0.07)	2.22 (0.28)
$\tilde{u}^2_{t-1}$	0.26 (0.06)	0.14 (0.07)	0.02 (0.04)
$\tilde{u}^2_{t-2}$	0.12 (0.06)	0.11 (0.04)	0.15 (0.05)
g_2	5.54 (1.26)	2.05 (0.33)	2.92 (0.61)
g_3	153.25 (23.06)	8.63 (1.57)	6.45 (1.50)
$d_{t-1} \cdot \tilde{u}^2_{t-1}$		0.38 (0.11)	0.32 (0.09)
p_{11}	0.99 (*)	0.98 (*)	0.98 (*)
p_{22}	0.94 (*)	0.60 (*)	0.61 (*)
p_{33}	0.96 (*)	0.00 (*)	0.00 (*)
$Erg.\quad P_1$	0.60	0.37	0.44
$Erg.\quad P_2$	0.15	0.45	0.40
$Erg.\quad P_3$	0.25	0.18	0.16
Log L	909.67	-1946.38	-2482.63
Distrib.	$N(\cdot)$	$N(\cdot)$	$N(\cdot)$

Notes:
(1) Standard Errors in parentheses.
(2) (*) Denotes that the probability-parameters in the model were changed to satisfy boundary constraints. Standard-error estimates are feasible but were not attempted.

Table 11: Forecast Accuracy of ARCH Models for the Return on the US$

Model	Loss Function			
	MSE	MAE	$[LE]^2$	$\lvert LE \rvert$
OLS Homocedastic	0.110979	0.127350	22.13701	4.102885
EGARCH (1,1) GED	0.099713	0.094078*	11.72607	2.500886
EGARCH (1,1) $N(\cdot)$	0.106856	0.100949	7.898725*	2.088073*
GARCH (1,1) $N(\cdot)$	0.104442	0.101210	9.653705	2.257588
TARCH (1,1) $N(\cdot)$	0.108278	0.101349	8.215606	2.124275
GARCH (1,1) $t(\cdot)$	0.123511	0.123607	9.613312	2.467701
SWARCH (2,4) $N(\cdot)$	0.845339	0.129706	18.81351	2.636824
SWARCH (2,4) $t(\cdot)$	0.162752	0.128938	9.453510	2.391157
SWARCH (3,2) $N(\cdot)$	0.097276*	0.096689	10.38927	2.432782

Notes: (1) $MSE = T^{-1}\sum_{t=1}^{T}\left\{\hat{u}_t^2 - \hat{s}_t^2\right\}^2$, $MAE = T^{-1}\sum_{t=1}^{T}\left|\hat{u}_t^2 - \hat{s}_t^2\right|$

$[LE]^2 = T^{-1}\sum_{t=1}^{T}\left\{\ln\left(\hat{u}_t^2\right) - \ln\left(\hat{s}_t^2\right)\right\}^2$, $|LE| = T^{-1}\sum_{t=1}^{T}\left|\ln\left(\hat{u}_t^2\right) - \ln\left(\hat{s}_t^2\right)\right|$

(2) * Denotes the best model.

Table 12: Forecast Accuracy of ARCH Models for the Return on the CBOND

Model	Loss Function			
	MSE	MAE	$[LE]^2$	$\lvert LE \rvert$
OLS Homocedastic	206.2718	4.581094	12.04767	2.608068
EGARCH (1,1) GED	192.1416	4.077298	11.09051	2.264638
EGARCH (1,1) $N(\cdot)$	173.8946*	3.768738*	9.832155	2.211236
GARCH (1,1) $N(\cdot)$	191.2182	4.191442	8.498950	2.138627*
TARCH (1,1) $N(\cdot)$	179.9914	3.952777	9.692357	2.209044
GARCH (1,1) $t(\cdot)$	192.8944	4.258412	8.430989*	2.157895
SWARCHL (2,3) $N(\cdot)$	184.0056	3.916024	8.440727	2.158818
SWARCHL (2,3) $t(\cdot)$	183.4115	3.865278	8.449593	2.157306
SWARCHL (3,2) $N(\cdot)$	179.1721	3.839258	8.586411	2.153293

Notes: (1) $MSE = T^{-1}\sum_{t=1}^{T}\left\{\hat{u}_t^2 - \hat{s}_t^2\right\}^2$, $MAE = T^{-1}\sum_{t=1}^{T}\left|\hat{u}_t^2 - \hat{s}_t^2\right|$

$[LE]^2 = T^{-1}\sum_{t=1}^{T}\left\{\ln\left(\hat{u}_t^2\right) - \ln\left(\hat{s}_t^2\right)\right\}^2$, $|LE| = T^{-1}\sum_{t=1}^{T}\left|\ln\left(\hat{u}_t^2\right) - \ln\left(\hat{s}_t^2\right)\right|$

(2) * Denotes the best model.

Table 13: Forecast Accuracy of ARCH Models for the Return on the COCOA

Model	Loss Function			
	MSE	MAE	$[LE]^2$	$\lvert LE \rvert$
OLS Homocedastic	60.29065	3.248362	12.25410	2.554517
EGARCH (1,1) GED	58.96771	3.064281*	85.07656	4.501383
EGARCH (1,1) $N(\cdot)$	50.80501*	3.070752	9.683910*	2.307081*
GARCH (1,1) $N(\cdot)$	51.08158	3.087768	11.88679	2.457249
TARCH (1,1) $N(\cdot)$	58.61194	3.147152	11.17255	2.408486
GARCH (1,1) $t(\cdot)$	51.84851	3.291344	15.30956	2.698542
SWARCH (2,4) $N(\cdot)$	52.84229	3.168131	17.49321	2.814531
SWARCH (2,2) $t(\cdot)$	51.59023	3.170375	18.33362	2.833525

Notes: (1) $MSE = T^{-1}\sum_{t=1}^{T}\left\{\hat{u}_t^2 - \hat{s}_t^2\right\}^2$, $MAE = T^{-1}\sum_{t=1}^{T}\left|\hat{u}_t^2 - \hat{s}_t^2\right|$

$[LE]^2 = T^{-1}\sum_{t=1}^{T}\left\{\ln\left(\hat{u}_t^2\right) - \ln\left(\hat{s}_t^2\right)\right\}^2$, $|LE| = T^{-1}\sum_{t=1}^{T}\left|\ln\left(\hat{u}_t^2\right) - \ln\left(\hat{s}_t^2\right)\right|$

(2) * Denotes the best model.

Table 14: Forecast Accuracy of ARCH Models for the Return on TELEBRAS

Model	Loss Function			
	MSE	MAE	$[LE]^2$	$\lvert LE \rvert$
OLS Homocedastic	1595.468	12.26207	9.962296	2.302542
EGARCH (1,1) GED	1412.165	10.31947	8.001585	1.985133
EGARCH (1,1) $N(\cdot)$	1360.067	9.924305*	7.512093	1.933137
GARCH (1,1) $N(\cdot)$	1503.709	10.99655	6.857071*	1.893813*
TARCH (1,1) $N(\cdot)$	1347.081*	10.03158	7.581257	1.938151
GARCH (1,1) $t(\cdot)$	1513.173	11.16956	6.916737	1.902861
SWARCHL(2,4) $N(\cdot)$	1387.448	10.27614	6.869221	1.898802
SWARCHL (2,4) $t(\cdot)$	1387.997	10.34076	6.912706	1.894535
SWARCHL(3,2) $N(\cdot)$	1367.935	10.09478	7.073707	1.910460

Notes: (1) $MSE = T^{-1}\sum_{t=1}^{T}\left\{\hat{u}_t^2 - \hat{s}_t^2\right\}^2$, $MAE = T^{-1}\sum_{t=1}^{T}\left|\hat{u}_t^2 - \hat{s}_t^2\right|$

$[LE]^2 = T^{-1}\sum_{t=1}^{T}\left\{\ln\left(\hat{u}_t^2\right) - \ln\left(\hat{s}_t^2\right)\right\}^2$, $|LE| = T^{-1}\sum_{t=1}^{T}\left|\ln\left(\hat{u}_t^2\right) - \ln\left(\hat{s}_t^2\right)\right|$

(3) * Denotes the best model.

Table 15: Goodness-of-Fit Measures for ARCH Estimates of the Return of the US$

Model	$Max\,[Log\ L]$	AIC	BIC
OLS Homocedastic	-157.6977	0.305482	0.310242
GARCH (1,1) $N(\cdot)$	872.31	-1.6714	-1.6524
EGARCH (1,1) $N(\cdot)$	876.79	-1.6781	-1.6543
GARCH (1,1) $t(\cdot)$	945.66	-1.8107	-1.7869
EGARCH (1,1) GED	931.75	-1.782	-1.7530
TARCH (1,1) $N(\cdot)$	882.56	-1.689	-1.1665
SWARCH (2,4) $N(\cdot)$	904.39	-1.7236	-1.6807
SWARCH (2,4) $t(\cdot)$	968.72*	-1.8455*	-1.7979*
SWARCH (3,2) $N(\cdot)$	909.67	-1.7318	-1.6842

Notes:
(1) * Denotes the best model.

Table 16: Goodness-of-Fit Measures for ARCH Estimates of the Return on the CBOND

Model	$Max\left[Log\ L\right]$	AIC	BIC
OLS Homocedastic	-2239.394	4.095785	4.100353
GARCH (1,1) $N(\cdot)$	-2001.36	3.6661	3.6844
EGARCH (1,1) $N(\cdot)$	-1981.85	3.6323	3.6551
GARCH (1,1) $t(\cdot)$	-1928.77*	3.5352*	3.5439*
EGARCH (1,1) GED	-1937.33	3.553	3.580
TARCH (1,1) $N(\cdot)$	-1980.54	3.630	3.653
SWARCH (2,3) $N(\cdot)$	-1944.22	3.5708	3.6119
SWARCH (2,3) $t(\cdot)$	-1943.26	3.5709	3.6165
SWARCH (3,2) $N(\cdot)$	-1946.38	3.5766	3.6222

Notes:
(1) * Denotes the best model.

Table 17: Goodness-of-Fit Measures for ARCH Estimates of the Return on COCOA

Model	$Max\left[Log\ L\right]$	AIC	BIC
OLS Homocedastic	-4219.340	3.819312	3.821891
GARCH (1,1) $N(\cdot)$	-4078.65	3.6947	3.7050
EGARCH (1,1) $N(\cdot)$	-4090.70	3.7065	3.7194
GARCH (1,1) $t(\cdot)$	-3910.23	3.5432	3.5561
EGARCH (1,1) GED	-3889.42*	3.5190*	3.5340*
TARCH (1,1) $N(\cdot)$	-4077.27	3.694	3.707
SWARCH (2,4) $N(\cdot)$	-3968.02	3.6000	3.6258
SWARCH (2,2) $t(\cdot)$	-3912.25	3.5468	3.5649

Notes:
(1) * Denotes the best model.

Table 18: Goodness-of-Fit Measures for ARCH Estimates of the Return on TELEBRAS

Model	$Max\left[Log\ L\right]$	AIC	BIC
OLS Homocedastic	-2691.018	5.181940	5.186701
GARCH (1,1) $N(\cdot)$	-2498.84	4.8178	4.8368
EGARCH (1,1) $N(\cdot)$	-2483.38	4.7899	4.8137
GARCH (1,1) $t(\cdot)$	-2485.67	4.7944	4.8034*
EGARCH (1,1) GED	-2488.63	4.802	4.831
TARCH (1,1) $N(\cdot)$	-2483.45	4.7901	4.8139
SWARCH (2,4) $N(\cdot)$	-2500.07	4.8298	4.8726
SWARCH (2,4) $t(\cdot)$	-2475.90*	4.7852*	4.8328
SWARCH (3,2) $N(\cdot)$	-2482.63	4.7981	4.8457

Notes:
(1) * Denotes the best model.